

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 2, 2004, 19:47:31; Search time 5.78967 Seconds

(without alignments)  
371,762 Million cell updates/sec

Title: US-10-054-873-3

Perfect score: 33

Sequence: 1 LGTPR 6

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: A\_Geneseq\_23Sep04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	6	AAV42857	AAV42857 Cleavable
2	33	100.0	15	ADM08409	ADM08409 Canine hm
3	33	100.0	15	ADM08322	ADM08322 Canine hm
4	33	100.0	107	AAV42860	AAV42860 hGH-mini-
5	33	100.0	135	AAV39387	AAV39387 Human pro
6	33	100.0	135	ADJ68719	ADJ68719 Human hea
7	33	100.0	150	AAV42861	AAV42861 Chimeric
8	33	100.0	388	AAO69171	AAO69171 Pseudomon
9	33	100.0	392	AAV42128	AAV42128 Human pol
10	33	100.0	405	ABG22253	ABG22253 Novel hum
11	33	100.0	585	AAV40342	AAV40342 Human pol
12	33	100.0	952	ADM04599	ADM04599 Human pro
13	33	100.0	1547	ABU96680	ABU96680 Human nuc
14	33	100.0	1562	ABP42703	ABP42703 Alpha -NA
15	33	100.0	2013	ABH47751	ABH47751 Listeria
16	33	100.0	2013	AAU32684	AAU32684 Protein e
17	33	100.0	7257	AAV58576	AAV58576 Sorangium
18	31	99.9	238	AAU27512	AAU27512 Human G-P
19	31	99.9	347	ABO29145	ABO29145 Mouse nov
20	30	90.9	448	ABP66801	ABP66801 Bifidobac
21	30	90.9	24	ABH39575	ABH39575 Peptide #
22	30	90.9	24	AAV33134	AAV33134 Peptide #
23	30	90.9	24	ABH24291	ABH24291 Protein #
24	30	90.9	24	AAV72803	AAV72803 Human bon
25	30	90.9	24	AAV60274	AAV60274 Human bra

26	30	50.9	24	ABG54611	ABG54611 Human liv
27	30	50.9	24	ABG42737	ABG42737 Human pep
28	30	50.9	100	AAU49418	AAU49418 Protiomb
29	30	50.9	100	ABW45937	ABW45937 Protiomb
30	30	50.9	108	AAO4045	AAO4045 Human sec
31	30	50.9	110	AAO2964	AAO2964 Human sec
32	30	50.9	114	AAH89914	AAH89914 C Glytam
33	30	50.9	203	AAV79780	AAV79780 Coryneb
34	30	50.9	209	AAV94054	AAV94054 Human pro
35	30	50.9	209	ADA03124	ADA03124 Nucleotid
36	30	50.9	222	AAU68588	AAU68588 Human nov
37	30	50.9	231	ABO00555	ABO00555 Novel hum
38	30	50.9	288	AAH67569	AAH67569 Amino aci
39	30	50.9	288	ABO07232	ABO07232 Human p53
40	30	50.9	298	AAV73880	AAV73880 Human pro
41	30	50.9	298	AAV43779	AAV43779 Human can
42	30	50.9	298	AAU23487	AAU23487 Novel hum
43	30	50.9	298	AAV25792	AAV25792 Human pro
44	30	50.9	301	ABH06045	ABH06045 Human NS
45	30	50.9	325	ABU96711	ABU96711 Human nuc

## ALIGNMENTS

RESULT 1  
AAV42857  
ID AAV42857 standard; peptide; 6 AA.

AC AAV42857;  
DT 19-JAN-2000 (first entry)  
DE Cleavable peptide linker for hGH-mini-proinsulin chimeric protein.

DE Linker: growth hormone; chaperone; intramolecular; insulin; precursor;  
KM folding; conformation; chimeric protein; cleavable; recombinant;  
KM production; yield.

OS Synthetic.

PN WO950302-A1.

PD 07-OCT-1999.

PF 31-MAR-1998; 98WO-CN000052.

PR 31-MAR-1998; 98WO-CN000052.

PA (TONG-) TONGHUA GAMTECH BIOTECHNOLOGY LTD.

PI Gan Z;

DR WPI; 1999-610839/52.

PT New chimeric proteins containing human growth hormone fragment, used particularly for the production of human insulin.

PS Claim 6; Page 29; 46pp; English.

This sequence represents a cleavable peptide linker which is a component of the chimeric proteins hGH-mini-proinsulin (AAV42860) and the chimeric protein given in AAV42861. These chimeric proteins additionally contain an N-terminal fragment of human growth hormone (hGH) and a human insulin precursor (AAV42859). The hGH portion of the chimeric protein acts as an intramolecular chaperone (IMC) for the insulin precursor, enabling it to fold correctly. The cleavable peptide linker has a C-terminal Arg residue (AAV42857) which enables the hGH portion of the chimeric protein to be removed after folding has taken place. Production of recombinant human insulin via an hGH-proinsulin chimeric protein can provide human insulin with correctly linked cysteine bridges with fewer necessary procedural steps, and hence resulting in a higher yield of human insulin. The IMC sequences not only protect insulin sequences from intracellular

degradation by a microorganism host, but also promote the folding of the fused insulin precursor, facilitate the solubility of the fusion protein, and decrease the intermolecular interactions among the fusion proteins, thus allowing folding of the fused insulin precursor at commercially useful high concentrations. The procedural steps of cyanogen bromide cleavage, oxidative sulphydrololysis and related purification steps can thus be eliminated, along with the use of high concentrations of mercaptan or the use of hydrophobic absorbent resins

XX Sequence 6 AA;

XX Query Match

Best Local Similarity 100.0%; Score 33; DB 2; Length 6;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LGTGR 6

Db 1 LGTGR 6

RESULT 2

ADM08409 standard; peptide; 15 AA.

XX ADM08409;

XX 20-MAY-2004 (first entry)

XX Canine immunoglobulin group 3 VL species framework 2 peptide 22.

XX canine; dog; heavy; immunoglobulin; antibody light chain variable domain;

XX antiallergic; allergy; IGE; gene therapy; group 3 species; VL framework;

XX FR2.

XX Canis familiaris.

XX WO2003060080-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US041362.

XX 21-DEC-2001; 2001US-0344874P.

XX (IDEX-) IDEXX LAB INC.

XX Krah ER, Guo H, Aiyappa A, Lawton R;

XX WPI; 2003-598521/56.

XX New canine heavy and light chain variable domain polypeptides, useful for

XX treating canine allergy.

XX Claim 40; Page 107; 130pp; English.

XX The invention relates to a novel canine heavy or light chain variable

XX domain polypeptide. The protein of the invention demonstrates

XX antiallergic activity and may be useful for treating canine allergy,

XX possibly via gene therapy. The current sequence is that of a canine

XX immunoglobulin light chain variable domain framework (FR) peptide of the

XX invention.

XX Sequence 15 AA;

XX Query Match 100.0%; Score 33; DB 7; Length 15;

XX Best Local Similarity 100.0%; Pred. No. 14;

XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LGTGR 6

Db 6 LGTGR 11

RESULT 3

ADM08322 standard; peptide; 15 AA.

XX ADM08322;

XX 20-MAY-2004 (first entry)

XX Canine immunoglobulin group 3 VL subgenus framework 2 peptide 8.

XX canine; dog; heavy; immunoglobulin; antibody light chain variable domain;

XX antiallergic; allergy; IGE; gene therapy; group 3 subgenus; VL framework;

XX FR2.

XX Canis familiaris.

XX WO2003060080-A2.

XX 24-JUL-2003.

XX 20-DEC-2002; 2002WO-US041362.

XX 21-DEC-2001; 2001US-0344874P.

XX (IDEX-) IDEXX LAB INC.

XX Krah ER, Guo H, Aiyappa A, Lawton R;

XX WPI; 2003-598521/56.

XX New canine heavy and light chain variable domain polypeptides, useful for

XX treating canine allergy.

XX Claim 39; Page 106; 130pp; English.

XX The invention relates to a novel canine heavy or light chain variable

XX domain polypeptide. The protein of the invention demonstrates

XX antiallergic activity and may be useful for treating canine allergy,

XX possibly via gene therapy. The current sequence is that of a canine

XX immunoglobulin light chain variable domain framework (FR) peptide of the

XX invention.

XX Sequence 15 AA;

XX Query Match 100.0%; Score 33; DB 7; Length 15;

XX Best Local Similarity 100.0%; Pred. No. 14;

XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LGTGR 6

Db 6 LGTGR 11

RESULT 4

AAV42860 standard; protein; 107 AA.

XX AAV42860;

XX 19-JAN-2000 (first entry)

XX hGH-mini-proinsulin chimeric protein.

XX Insulin; precursor; growth hormone; chaperone; intramolecular; folding;

XX conformation; chimeric protein; cleavable; recombinant; production;

XX yield.

XX Synthetic.

XX Homo sapiens.

XX WO950302-A1.

XX 07-OCT-1999.

XX PF 31-MAR-1998; 98WO-CN000052.  
 XX PR 31-MAR-1998; 98WO-CN000052.  
 XX PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.  
 XX PI Gan Z;  
 XX DR WPI; 1999-610839/52.  
 XX PT New chimeric proteins containing human growth hormone fragment, used  
 XX PT particularly for the production of human insulin.  
 XX PS Claim 13; Page 30; 46pp; English.

CC This sequence represents a chimeric protein, hGH-mini-proinsulin. This  
 CC chimeric protein contains an N-terminal fragment of human growth hormone  
 CC (hGH) of the sequence given in AA42855, a cleavable peptide linker  
 CC (AA42857), and a human insulin precursor comprising insulin A and B  
 CC chains (AA42859). The hGH portion of the chimeric protein acts as an  
 CC intramolecular chaperone (IMC) for the insulin precursor, enabling it to  
 CC fold correctly. The cleavable peptide linker has a C-terminal Arg residue  
 CC which enables the hGH portion of the chimeric protein to be removed after  
 CC folding has taken place. Production of recombinant human insulin via an  
 CC hGH-proinsulin chimeric protein can provide human insulin with correctly  
 CC linked cysteine bridges with fewer necessary procedural steps, and hence  
 CC resulting in a higher yield of human insulin. The IMC sequences not only  
 CC protect insulin sequences from intracellular degradation by a  
 CC microorganism host, but also promote the folding of the fused insulin  
 CC precursor, facilitate the solubility of the fusion protein and decrease  
 CC the intermolecular interactions among the fusion proteins, thus allowing  
 CC folding of the fused insulin precursor at commercially useful high  
 CC concentrations. The procedural steps of cyanogen bromide cleavage,  
 CC oxidative sulphyloysis and related purification steps can thus be  
 CC eliminated, along with the use of high concentrations of mercaptan or the  
 CC use of hydrophobic absorbent resins

XX SQ Sequence 107 AA;

Query Match 100.0%; Score 33; DB 2; Length 107;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
 |||||  
 Db 50 LGTGR 55

RESULT 5  
 AAB93957  
 ID AAB93957 standard; protein; 135 AA.  
 XX AAB93957;  
 AC AAB93957;  
 XX 26-JUN-2001 (first entry)  
 DT 26-JUN-2001 (first entry)  
 XX Human protein sequence SEQ ID NO:14002.  
 DE Human protein sequence SEQ ID NO:14002.  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 OS Homo sapiens.  
 XX Homo sapiens.  
 XX EP1074617-A2.  
 PN EP1074617-A2.  
 XX 07-FEB-2001.  
 PD 07-FEB-2001.  
 XX 28-JUL-2000; 2000EP-00116126.  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300252.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00185767.

PR 09-JUN-2000; 2000JP-00241899.  
 XX (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX DR WPI; 2001-318749/34.  
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 XX PT length cDNAs defined in the specification, and for the detection and/or  
 XX PT diagnosis of the abnormality of the proteins encoded by the full-length  
 XX PT cDNAs.

XX PS Claim 8; SEQ ID NO 14002; 2537pp + Sequence listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention

XX SQ Sequence 135 AA;

Query Match 100.0%; Score 33; DB 4; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
 |||||  
 Db 5 LGTGR 10

RESULT 6  
 ADJ69719  
 ID ADJ69719 standard; protein; 135 AA.  
 XX ADJ69719;  
 AC ADJ69719;  
 XX 06-MAY-2004 (first entry)  
 DT 06-MAY-2004 (first entry)  
 XX Human heat mitochondrial protein as a therapeutic target SeqID1525.  
 DE Human heat mitochondrial protein as a therapeutic target SeqID1525.  
 XX Mitochondrial; human; screening assay; diabetes mellitus;  
 KM Mitochondrial; human; screening assay; diabetes mellitus;  
 KM Huntington's disease; osteoarthritis;  
 KM Leber's hereditary optic neuropathy; LHON;  
 KM mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
 KM myoclonic epilepsy ragged red fibre syndrome; MERRF; Cancer;  
 KM neuroprotective; noctropic; antidiabetic; anticonvulsant; antihypertic;  
 KM osteopathic; ophthalmological; cytosstatic.  
 XX Homo sapiens.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO2003087768-A2.  
 PN WO2003087768-A2.  
 XX 23-OCT-2003.

XX	04-APR-2003; 2003MO-US010870.
XX	12-APR-2002; 2002US-0372843P.
PR	17-JUN-2002; 2002US-0389987F.
PR	20-SEP-2002; 2002US-0412418P.
XX	(MITO-) MITOKOR.
PA	(BUCK-) BUCK INST AGE RES.
XX	Gnosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI	Marnock DE;
DR	WPI; 2003-845369/78.
XX	
PT	Identifying a mitochondrial target for drug screening assays and for
PT	treating diseases associated with altered mitochondrial function,
PT	comprises detecting a modified polypeptide in a sample and correlating
PT	with the disease.
XX	
PS	Claim 1; SEQ ID NO 1525; 180bp; English.
CC	This invention relates to novel mitochondrial targets that can be used
CC	for therapeutic intervention in treating a disease associated with
CC	altered mitochondrial function. Specifically, it refers to a method for
CC	identifying proteins of the human heart mitochondrial proteome that are
CC	useful for drug screening assays, as well as therapeutic targets. The
CC	present invention describes a method for identifying such proteins that
CC	can be used in the treatment of various diseases associated with altered
CC	mitochondrial function including diabetes mellitus, Huntington's disease,
CC	osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC	encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC	ragged red fibre syndrome (MERFF) or cancer. Accordingly, these
CC	compositions have neuroprotective, nootropic, antidiabetic,
CC	anticoagulant, antiarthritic, osteopathic, ophthalmological and
CC	cytoskeletal activities. This polypeptide sequence is a human heart
CC	mitochondrial protein of the invention.
XX	
SQ	Sequence 135 AA;
Query Match	100.0%; Score 33; DB 7; Length 135;
Best Local Similarity	100.0%; Pred. No. 1.3e+02;
Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 LGTGR 6 
DB	5 LGTGR 10 
RESULT 7	
ID	AAV42861 standard; protein; 150 AA.
AC	AAV42861;
DX	
DJ	19-JAN-2000 (first entry)
DE	Chimeric protein, SEQ ID 7.
XX	Inulin; precursor; growth hormone; chaperone; intramolecular; folding;
KW	conformation; chimeric protein; cleavable; recombinant; production;
KW	yield.
XX	Synthetic.
OS	Homo sapiens.
XX	
PN	WO950302-AI.
XX	
PD	07-OCT-1999.
DF	31-MAR-1998; 98WO-CN000052.
XX	
PR	31-MAR-1998; 98WO-CN000052.

PA	(TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
XX	
PI	Gan 2;
XX	
DR	WPI; 1999-610839/52.
XX	
XX	New chimeric proteins containing human growth hormone fragment, used
PT	particularly for the production of human insulin.
XX	
PS	Claim 14; Page 30-31; 46pp; English.
XX	
CC	This sequence represents a chimeric protein, which contains an N-terminal
CC	fragment of human growth hormone (hGH) of the sequence given in AA142856,
CC	a cleavable peptide linker (AA142857), and a human insulin precursor
CC	comprising insulin A and B chains (AA142859). The hGH portion of the
CC	chimeric protein acts as an intramolecular chaperone (IMC) for the
CC	insulin precursor, enabling it to fold correctly. The cleavable peptide
CC	linker has a C-terminal Arg residue which enables the hGH portion of the
CC	chimeric protein to be removed after folding has taken place. Production
CC	of recombinant human insulin via an hGH-proinsulin chimeric protein can
CC	provide human insulin with correctly linked cysteine bridges with fewer
CC	necessary procedural steps, and hence resulting in a higher yield of
CC	human insulin. The IMC sequences not only protect insulin sequences from
CC	intracellular degradation by a microorganism host, but also promote the
CC	folding of the fused insulin precursor, facilitate the solubility of the
CC	fusion protein and decrease the intermolecular interactions among the
CC	fusion proteins, thus allowing folding of the fused insulin precursor at
CC	commercially useful high concentrations. The procedural steps of cyanogen
CC	bromide cleavage, oxidative subunitolysis and related purification steps
CC	can thus be eliminated, along with the use of high concentrations of
CC	mercaptan or the use of hydrophobic absorbent resins
XX	
SQ	Sequence 150 AA;
XX	
Query March	100.0%; Score 33; DB 2; Length 150;
Best local Similarity	100.0%; Pred. No. 1,4e+02;
Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	1 LGTGR 6
DB	93 LGTGR 98
XX	
RESULT 8	
AB069171	
ID	AB069171 standard; protein; 388 AA.
XX	
AC	AB069171;
XX	
DT	29-JUL-2004 (first entry)
XX	
DR	Pseudomonas aeruginosa polypeptide #1346.
XX	
KM	Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
XX	
OS	Pseudomonas aeruginosa.
XX	
PN	US6551795-B1.
XX	
PD	22-APR-2003.
XX	
PF	18-FEB-1999; 99US-00252991.
XX	
PR	18-FEB-1998; 98US-0074788P.
XX	
PR	27-JUL-1998; 98US-0094190P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Rubinfeld M, Nolling J, Delonghery C, Bush D;
XX	
XX	WPI; 2003-615309/58.
DR	N-PSDB; ABD02742.

XX Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 XX  
 PS Disclosure: SEQ ID NO 17917; 455pp; English.  
 XX  
 CC The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
 CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
 CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
 CC of *Pseudomonas* species using biochip technology. Sequences AB067826-  
 CC AB084396 represent *P. aeruginosa* polypeptides of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
 CC  
 XX Sequence 388 AA;  
 SQ  
 Query Match 100.0%; Score 33; DB 7; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTGR 6  
 DB 179 LGTGR 184  
 RESULT 9  
 AAM42128  
 ID AAM42128 standard; protein; 392 AA.  
 XX  
 AC AAM42128;  
 DT 22-OCT-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 7059.  
 XX  
 KW Human; nocrotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-UTL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US034263.  
 XX  
 PR 23-DEC-1999; 99US-00471275.  
 PR 21-JAN-2000; 2000US-00488725.  
 PR 25-APR-2000; 2000US-00552317.  
 PR 20-JUN-2000; 2000US-00558042.  
 PR 19-JUL-2000; 2000US-00620332.  
 PR 03-AUG-2000; 2000US-00633450.  
 PR 14-SEP-2000; 2000US-00662191.  
 PR 19-OCT-2000; 2000US-00693036.  
 PR 29-NOV-2000; 2000US-00727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
 PI Zhou P, Goodrich R, Dirmnac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR N-PSDB; AA161284.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 XX  
 PS Example 2; SEQ ID NO 7059; 10078pp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and the  
 CC encoded polypeptides (AAM8642-AAM42213) with nocrotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification  
 CC  
 XX Sequence 392 AA;  
 SQ  
 Query Match 100.0%; Score 33; DB 4; Length 392;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTGR 6  
 DB 17 LGTGR 22  
 RESULT 10  
 ABG22253  
 ID ABG22253 standard; protein; 405 AA.  
 XX  
 AC ABG22253;  
 DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #22244.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Dirmnac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS86440.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX Claim 20; SEQ ID NO 52612; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences: (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping;

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPPO at

CC ftp.wippo.int/pub/published\_pct\_sequences

XX

XX Sequence 405 AA;

XX

XX Query Match 100.0%; Score 33; DB 4; Length 405;

XX Best Local Similarity 100.0%; Pred. No. 3.7e+02;

XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 1 LGTGR 6

XX 87 LGTGR 92

XX

XX Db

XX

XX RESULT 11

XX AAM40342

XX ID AAM40342 standard; protein; 585 AA.

XX

XX AAM40342;

XX

XX 22-OCT-2001 (first entry)

XX

XX Human polypeptide SEQ ID NO 3487.

XX

XX Human; nocrotropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX peripheral nervous system; neuropathy; central nervous system; CNS;

XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX chemokine; chromolytic; drug screening; arthritis; inflammation;

XX leukaemia.

XX

XX Homo sapiens.

XX

XX WO200153312-A1.

XX

XX 26-JUL-2001.

XX

XX 26-DEC-2000; 2000WO-US034263.

XX

XX 23-DEC-1999; 99US-00471275.

XX

XX 21-JAN-2000; 2000US-0048725.

XX

XX 25-APR-2000; 2000US-0052317.

XX

XX 20-JUN-2000; 2000US-00598042.

XX

XX 19-JUL-2000; 2000US-00620312.

XX

XX 03-AUG-2000; 2000US-00653450.

XX

XX 14-SEP-2000; 2000US-00662191.

XX

XX 19-OCT-2000; 2000US-00693036.

XX

XX 29-NOV-2000; 2000US-00727344.

XX

XX (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

PI Zhou P, Goodrich R, Drmanac RT;

XX

XX WPI; 2001-442253/47.

XX

XX N-PSDB; AA159498.

XX

XX Novel nucleic acids and polypeptides, useful for treating disorders such

XX as central nervous system injuries.

XX

XX Example 6; SEQ ID NO 3487; 10078bp; English.

XX

XX The invention relates to human nucleic acids (AA157798-AA161369) and the

XX encoded polypeptides (AAM38642-AAM42213) with nocrotropic

XX immunosuppressant and cytostatic activity. The polynucleotides are useful

XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX of the invention may be used to treat diseases of the peripheral nervous

XX system, such as peripheral nervous injuries, peripheral neuropathy and

XX localised neuropathies and central nervous system diseases, such as

XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX utilisation of the activities such as: immune system suppression,

XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX assays for receptor activity, arthritis and inflammation, leukaemia and

XX C.N.S disorders. Note: The sequence data for this patent did not form

XX part of the printed specification

XX

XX Sequence 585 AA;

XX

XX Query Match 100.0%; Score 33; DB 4; Length 585;

XX Best Local Similarity 100.0%; Pred. No. 5.3e+02;

XX Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 1 LGTGR 6

XX 5 LGTGR 10

XX

XX Db

XX

XX RESULT 12

XX ADM04599

XX ID ADM04599 standard; protein; 952 AA.

XX

XX ADM04599;

XX

XX 20-MAY-2004 (first entry)

XX

XX Human protein of the invention SEQ ID NO:3284.

XX

XX Human; gene therapy; diagnostic marker; pharmaceutical.

XX

XX Homo sapiens.

XX

XX EP1347046-A1.

XX

XX 24-SEP-2003.

XX

XX 12-APR-2002; 2002EP-00008400.

XX

XX 22-MAR-2002; 2002JP-00137785.

XX

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX

XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

XX Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

XX Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;

XX

XX WPI; 2003-723558/69.

XX

XX N-PSDB; ADM02156.

XX

XX New polynucleotides and polypeptides are useful in gene therapy, for

XX developing a diagnostic marker or medicines for regulating their

XX expression and activity, or as a target of gene therapy.

XX PS Claim 1; SEQ ID NO 3284; 305bp; English.  
 XX CC The invention relates to a novel human polynucleotide and the encoded  
 CC polypeptide. A polynucleotide of the invention may have a use in gene  
 CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful  
 CC as a primer for synthesizing the polynucleotide or as a probe for  
 CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are  
 CC useful in gene therapy, for developing a diagnostic marker or medicines  
 CC for regulating their expression and activity, or as a target of gene  
 CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides  
 CC are useful as pharmaceutical agents. The present sequence represents a  
 CC protein sequence of the invention.  
 XX SQ Sequence 952 AA;  
 Query Match 100.0%; Score 33; DB 7; Length 952;  
 Best Local Similarity 100.0%; Pred. No. 8.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTSPR 6  
 Db 634 LGTSPR 639  
 RESULT 13  
 ABU96680  
 ID ABU96680 standard; protein; 1547 AA.  
 XX AC ABU96680;  
 XX DT 25-JUL-2003 (first entry)  
 XX DE Human nucleic acid-associated protein (NMAP) #9.  
 XX KW Human; nucleic acid-associated protein; cytosolic; antiarteriosclerotic;  
 KW antiapoptotic; neurotrophic; neuroprotective; cerebroprotective; anti-HIV;  
 KW antiallergic; antiinflammatory; thyromimetic; gene therapy;  
 KW cell proliferative disorder; cancer; atherosclerosis;  
 KW neurological disorder; epilepsy; Huntington's disease; stroke;  
 KW immune disorder; inflammatory disorder; AIDS; allergy;  
 KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;  
 KW protein-protein interaction; drug-target interaction;  
 KW gene expression profile.  
 XX OS Homo sapiens.  
 XX PN WO2003023003-A2.  
 XX XX 20-MAR-2003.  
 XX PF 05-SEP-2002; 2002WO-US028540.  
 XX PR 07-SEP-2001; 2001US-0317792P.  
 PR 07-SEP-2001; 2001US-0317912P.  
 PR 14-SEP-2001; 2001US-0322270P.  
 PR 21-SEP-2001; 2001US-0324040P.  
 PR 28-SEP-2001; 2001US-0326732P.  
 PR 19-OCT-2001; 2001US-0346716P.  
 PR 25-JAN-2002; 2002US-0351749P.  
 PR 22-FEB-2002; 2002US-0359498P.  
 XX PA (INCYTE GENOMICS INC.  
 XX Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe J;  
 PI Becha SD, Richardson TW, Lee EA, Sprague WM, Emerling BM;  
 PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li UX;  
 PI Hafalia AD, Sanjanwala B, Marguis JP, Gorvad AE, Lee SY, Ison CH;  
 PI Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebajjadian Y, Shah P;  
 PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kadle AB;  
 PI Burford N, Ramkumar J;  
 XX DR WPI; 2003-313243/30.

DR N-PSDB; ACA98928.  
 XX PT New human nucleic acid associated proteins (NMAP), useful for diagnosing,  
 PT treating and preventing diseases or conditions associated with the  
 PT aberrant NMAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
 PT infections.  
 XX PS Claim 1; Page 243-247; 345bp; English.  
 XX SQ The invention describes a novel human isolated nucleic acid-associated  
 CC polypeptide (NMAP). The polypeptides and polynucleotides are useful in  
 CC diagnosing, treating and preventing diseases or conditions associated  
 CC with the decreased expression or overexpression of NMAP, such as cell  
 CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.  
 CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,  
 CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)  
 CC disorders, or infections. These are also useful in assessing the effects  
 CC of exogenous compounds on the expression of nucleic acid and amino acid  
 CC sequences of NMAP. The NMAP or its fragments are useful in screening  
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,  
 CC or in altering the expression of the target polynucleotide and compounds  
 CC that specifically bind to or modulate the activity of the polypeptide.  
 CC The microarray is useful in monitoring or measuring protein-protein  
 CC interactions, drug-target interactions, and gene expression profiles.  
 CC This is the amino acid sequence of a novel human nucleic acid-associated  
 CC protein (NMAP)  
 XX SQ Sequence 1547 AA;  
 Query Match 100.0%; Score 33; DB 6; Length 1547;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTSPR 6  
 Db 1276 LGTSPR 1281  
 RESULT 14  
 ABP43703  
 ID ABP43703 standard; protein; 1562 AA.  
 XX AC ABP43703;  
 XX DT 26-FEB-2003 (first entry)  
 XX DE Alpha -NAC muscle specific form gp220.  
 XX KW Neuroprotective; immunomodulator; cancer; chromosome 7q31-q32;  
 KW cytosolic; anti-inflammatory; gene therapy; nutritional supplement;  
 KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;  
 KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;  
 KW vulnery.  
 XX OS Mus musculus.  
 XX PN WO200231111-A2.  
 XX XX 18-APR-2002.  
 XX PF 11-OCT-2001; 2001WO-US027760.  
 XX PR 12-OCT-2000; 2000US-0067527.  
 XX PA (HYSEQ INC.  
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehman T, Dmanac RT;  
 PI WPI; 2002-426278/45.  
 XX DR N-PSDB; ABQ60947.  
 XX PT New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating  
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or  
 PT inflammation.  
 XX  
 XX  
 XX Claim 20; SEQ ID # 606; 357bp + Sequence Listing; English.  
 CC The invention relates to 446 newly isolated polynucleotide sequences. The  
 CC activity of polynucleotides of the invention may be described as,  
 CC inflammatory, neuroprotective, immunomodulatory, cytoskeletal and anti-  
 CC inflammatory. Compositions comprising nucleic acids of the invention are  
 CC useful for treating a mammalian subject, or as nutritional sources or  
 CC supplements. These are useful in gene therapy, particularly for treating  
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,  
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or  
 CC inflammation. The nucleic acids and polypeptides are also useful in  
 CC diagnostic and research methods. The sequences given in records ABP43544-  
 CC ABP43589 represent polypeptides encoded by polynucleotides of the  
 CC invention. NOTE: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences  
 XX  
 SQ Sequence 1562 AA;  
 Query Match 100.0%; Score 33; DB 5; Length 1562;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTSPR 6  
 Db 1291 LGTSPR 1296  
 RESULT 15  
 ID ABB47751  
 XX ABB47751 standard; protein; 2013 AA.  
 AC ABB47751;  
 XX  
 XX 05-FEB-2002 (first entry)  
 DT  
 XX  
 DE Listeria monocytogenes protein #455.  
 XX  
 XX Listeria monocytogenes protein #455.  
 KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM Vitamin B12; bacterial infection; disease.  
 XX  
 XX Listeria monocytogenes.  
 OS  
 XX  
 XX WO200177335-A2.  
 PN  
 XX  
 XX 18-OCT-2001.  
 PD  
 XX  
 XX 11-APR-2001; 2001WO-FR001118.  
 PF  
 XX  
 XX 11-APR-2000; 2000FR-00004629.  
 PR  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA  
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Feizi H, Dehoux P,  
 PI Dussangot O, Cheonani F, Nedjari H, Glaser P, Kunst F, Cossart P,  
 PI Daniels J, Goebel W, Kieft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L,  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
 PI Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J,  
 PI Rose M, Voss H;  
 XX  
 XX WPI/ 2002-010914/01.  
 DR  
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and related  
 PT polypeptides.  
 XX  
 PS Claim 6; SEQ ID NO 456; 192bp; French.

XX  
 CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of L. monocytogenes and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccine compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms. NOTE: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pcl\_sequences  
 XX  
 SQ Sequence 2013 AA;  
 Query Match 100.0%; Score 33; DB 5; Length 2013;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTSPR 6  
 Db 692 LGTSPR 697  
 Search completed: November 2, 2004, 20:11:42  
 Job time : 8.78967 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 2, 2004, 20:02:41 ; Search time 1.46125 Seconds

(without alignments)  
272.306 Million cell updates/sec

Title: US-10-054-873-3

Perfect score: 33

Sequence: 1 LGTGR 6

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 478139 seqs, 6631800 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/1/1aa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	33	100.0	388	4	US-09-252-991A-17917	Sequence 17917, A
2	33	100.0	7257	3	US-09-335-409-5	Sequence 5, Appl
3	33	100.0	7257	3	US-09-568-102-5	Sequence 5, Appl
4	33	100.0	7257	3	US-09-567-969-5	Sequence 5, Appl
5	33	100.0	7257	3	US-09-568-480-5	Sequence 5, Appl
6	33	100.0	7257	3	US-09-568-486-5	Sequence 5, Appl
7	33	100.0	7257	3	US-09-567-899-5	Sequence 5, Appl
8	33	100.0	108	4	US-09-513-999C-8126	Sequence 8126, Ap
9	33	90.9	124	4	US-09-621-976-6823	Sequence 6823, Ap
10	33	90.9	455	4	US-09-470-767-46061	Sequence 46061, A
11	33	90.9	684	4	US-09-252-991A-32139	Sequence 32139, A
12	33	90.9	17	3	US-09-461-697-28	Sequence 24802, A
13	33	87.9	20	3	US-08-724-4668-9	Sequence 28, Appl
14	33	87.9	20	3	US-08-682-164D-9	Sequence 9, Appl
15	33	87.9	58	3	US-08-827-219-39	Sequence 9, Appl
16	33	87.9	84	4	US-09-583-110-5115	Sequence 39, Appl
17	33	87.9	143	4	US-09-252-991A-30955	Sequence 5115, Ap
18	33	87.9	153	4	US-09-621-976-4040	Sequence 30955, Ap
19	33	87.9	176	4	US-09-252-991A-18623	Sequence 4040, Ap
20	33	87.9	168	4	US-09-252-991A-24281	Sequence 18623, A
21	33	87.9	176	4	US-09-248-796A-17378	Sequence 24281, A
22	33	87.9	198	4	US-09-252-991A-19838	Sequence 17378, A
23	33	87.9	217	4	US-09-252-991A-21872	Sequence 19838, A
24	33	87.9	248	4	US-09-252-991A-21872	Sequence 21872, A
25	33	87.9	259	4	US-09-436-434-2	Sequence 2, Appl
26	33	87.9	259	4	US-09-252-991A-17461	Sequence 17461, A
27	33	87.9	259	4	US-09-252-991A-17461	Sequence 17461, A

28	29	87.9	274	4	US-09-252-991A-17147	Sequence 17147, A
29	29	87.9	293	4	US-09-252-991A-18793	Sequence 18793, A
30	29	87.9	305	4	US-09-710-279-1362	Sequence 1362, Ap
31	29	87.9	313	4	US-09-252-991A-23631	Sequence 23631, A
32	29	87.9	313	4	US-09-252-991A-23810	Sequence 23810, A
33	29	87.9	345	4	US-09-107-532A-4268	Sequence 4268, Ap
34	29	87.9	357	4	US-09-252-991A-28380	Sequence 28380, A
35	29	87.9	365	4	US-09-134-000C-4369	Sequence 4369, Ap
36	29	87.9	408	4	US-09-252-991A-20095	Sequence 20095, Ap
37	29	87.9	409	4	US-09-252-991A-23414	Sequence 23414, A
38	29	87.9	412	4	US-09-355-912A-5	Sequence 5, Appl
39	29	87.9	430	4	US-09-252-991A-32661	Sequence 32661, A
40	29	87.9	460	4	US-09-198-452A-1085	Sequence 1085, Ap
41	29	87.9	474	4	US-09-252-991A-29391	Sequence 29391, A
42	29	87.9	480	4	US-09-252-991A-24889	Sequence 24889, A
43	29	87.9	507	4	US-09-248-796A-15540	Sequence 15540, A
44	29	87.9	518	4	US-09-252-991A-23604	Sequence 23604, A
45	29	87.9	550	4	US-09-252-991A-21295	Sequence 21295, A

## ALIGNMENTS

```
RESULT 1
US-09-252-991A-17917
Sequence 17917, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17917
LENGTH: 388
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17917

Query Match      100.0%; Score 33; DB 4; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LGTGR 6
Db      179 LGTGR 184

RESULT 2
US-09-335-409-5
Sequence 5, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schnapp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOCHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 7257
```

TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-335-409-5

Query Match  
Best Local Similarity 100.0%; Score 33; DB 3; Length 7257;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 1041 LGTGR 1046

RESULT 3

US-09-568-102-5  
Sequence 5, Application US/09568102  
Patent No. 6346404

GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,102  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-568-102-5

Query Match  
Best Local Similarity 100.0%; Score 33; DB 3; Length 7257;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 1041 LGTGR 1046

RESULT 4

US-09-567-969-5  
Sequence 5, Application US/09567969  
Patent No. 6355457

GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/567,969  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-567-969-5

Query Match 100.0%; Score 33; DB 3; Length 7257;

Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 1041 LGTGR 1046

RESULT 5

US-09-568-480-5  
Sequence 5, Application US/09568480  
Patent No. 6355458

GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,480  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-568-480-5

Query Match  
Best Local Similarity 100.0%; Score 33; DB 3; Length 7257;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 1041 LGTGR 1046

RESULT 6

US-09-568-486-5  
Sequence 5, Application US/09568486  
Patent No. 6355459

GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-30582A  
CURRENT APPLICATION NUMBER: US/09/568,486  
CURRENT FILING DATE: 2000-05-10  
PRIOR APPLICATION NUMBER: 09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-568-486-5

Query Match  
Best Local Similarity 100.0%; Score 33; DB 3; Length 7257;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||

DB 1041 LGTGR 1046

```
RESULT 7
US-09-568-472-5
; Sequence 5, Application US/09568472
; Patent No. 6356719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Molnar, James
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5
```

```
Query Match          100.0%; Score 33; DB 3; Length 7257;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 LGTGR 6  
DB 1041 LGTGR 1046

```
RESULT 8
US-09-567-899-5
; Sequence 5, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligot, James
; APPLICANT: Molnar, James
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-5
```

```
Query Match          100.0%; Score 33; DB 3; Length 7257;
Best Local Similarity 100.0%; Pred. No. 2e+03;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 LGTGR 6  
DB 1041 LGTGR 1046RESULT 9  
US-09-513-999C-8126

; Sequence 8126, Application US/09513993C

```
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59, US2, REG
; CURRENT APPLICATION NUMBER: US/09/513,993C
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent .pm
; SEQ ID NO 8126
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 101
; OTHER INFORMATION: Xaa=ile or Met
US-09-513-999C-8126
```

```
Query Match          90.9%; Score 30; DB 4; Length 108;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 LGTGR 6  
DB 36 VGTGR 41

```
RESULT 10
US-09-621-976-6823
; Sequence 6823, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent .pm
; SEQ ID NO 6823
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-6823
```

```
Query Match          90.9%; Score 30; DB 4; Length 124;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 LGTGR 6  
DB 36 VGTGR 41

```
RESULT 11
US-09-270-767-46061
; Sequence 46061, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-034
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
```

NUMBER OF SEQ ID NOS: 62517  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 46061  
LENGTH: 455  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-270-767-46061

Query Match 90.9%; Score 30; DB 4; Length 455;  
Best Local Similarity 83.3%; Pred. No. 5.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 1.07 LGTGR 112

RESULT 12  
US-09-252-991A-32139  
Sequence 32139, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 31142  
SEQ ID NO 32139  
LENGTH: 542  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32139

Query Match 90.9%; Score 30; DB 4; Length 542;  
Best Local Similarity 83.3%; Pred. No. 6.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 261 VGTGR 266

RESULT 13  
US-09-252-991A-24802  
Sequence 24802, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 24802  
LENGTH: 684  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24802

Query Match 90.9%; Score 30; DB 4; Length 684;  
Best Local Similarity 83.3%; Pred. No. 8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 274 VGTGR 279

RESULT 14  
US-09-461-697-28  
Sequence 28, Application US/09461697  
Patent No. 627974  
GENERAL INFORMATION:  
APPLICANT: COGENT NEUROSCIENCE, Inc.  
APPLICANT: Lo, Donald C.  
APPLICANT: Barney, Shawn  
APPLICANT: Thomas, Mary Beth  
APPLICANT: Portbury, Stuart D.  
APPLICANT: Putnam, Kasuri  
APPLICANT: Katz, Lawrence C.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING  
TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING  
FILE REFERENCE: 10001-003-999  
CURRENT APPLICATION NUMBER: US/09/461,697  
CURRENT FILING DATE: 1999-12-14  
NUMBER OF SEQ ID NOS: 466  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 17  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-461-697-28

Query Match 87.9%; Score 29; DB 3; Length 17;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTGR 6  
DB 11 GTGR 15

RESULT 15  
US-08-724-466B-9  
Sequence 9, Application US/08724466B  
Patent No. 6063606  
GENERAL INFORMATION:  
APPLICANT: Petkovich, P. Martin, White, Jay A.  
APPLICANT: Beckett, Barbara R., Jones, Glenville  
TITLE OF INVENTION: Retinoid Metabolizing Protein  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Blake, Cassels & Graydon  
STREET: Box 25, Commerce Court West  
CITY: Toronto  
ZIP: M5L 1A9  
COUNTRY: Canada  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3 1/2 inch, 1.4 Mb storage  
COMPUTER: COMPAQ, IBM PC compatible  
OPERATING SYSTEM: MS-DOS 5.1  
SOFTWARE: WORD PERFECT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,466B  
FILING DATE: October 1, 1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/667,546  
FILING DATE: June 21, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunt, John C.  
REGISTRATION NUMBER: 36,424  
REFERENCE/DOCKET NUMBER: 50767/00004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 863-4344

Sat Nov 6 18:59:22 2004

us-10-054-873-3.ra1

Page 5

TELEFAX: (416) 863-2653  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-724-466B-9

Query Match 87.9%; Score 29; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 GTGPR 6  
|||  
Db 3 GTGPR 7

Search completed: November 2, 2004, 20:24:33  
Job time : 2.46125 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: November 2, 2004, 19:59:41 ; Search time 1.10701 Seconds  
(without alignments)  
521.495 Million cell updates/sec

Title: US-10-054-873-3

Perfect score: 33

Sequence: 1 LGTGR 6

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR.79.\*  
2: PIR1.\*  
3: PIR2.\*  
4: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	100.0	68	2 AF2809	hypothetical prote
2	33	100.0	931	2 T49710	related to glucan
3	33	100.0	2013	2 AD1128	probable peptidogl
4	33	100.0	2013	2 A11469	probable peptidogl
5	31	93.9	188	2 C87341	conserved hypotet
6	30	90.9	288	2 T44603	hypothetical prote
7	30	90.9	329	2 C84847	probable guanylate
8	30	90.9	387	2 T50675	guanylate kinase (
9	30	90.9	391	2 T46191	guanylate kinase-1
10	30	90.9	432	2 S49980	glutamate-5-semial
11	30	90.9	811	2 A41054	fasciclin II, tran
12	30	90.9	873	2 B41054	fasciclin II, PI-11
13	30	90.9	1002	2 A36691	Ca2+-transporting
14	29	87.9	85	2 G70824	hypothetical prote
15	29	87.9	90	2 B95105	conserved hypotet
16	29	87.9	100	2 B97973	hypothetical prote
17	29	87.9	136	2 T36624	hypothetical prote
18	29	87.9	169	2 A64330	hypothetical prote
19	29	87.9	212	2 T44591	hypothetical prote
20	29	87.9	219	1 T1BPR2	tail fiber protein
21	29	87.9	226	2 S27759	maturation-associat
22	29	87.9	232	1 MMBR13	25.5K membrane pro
23	29	87.9	238	2 T40820	proline-rich prote
24	29	87.9	243	2 S27758	maturation-associat
25	29	87.9	243	2 A33263	exsp protein limpo
26	29	87.9	256	2 T11669	protein CPD22, dr
27	29	87.9	259	2 T37915	hypothetical prote
28	29	87.9	260	2 A36949	28.5K basic DNA-BI
29	29	87.9	302	2 T15936	hypothetical prote

30	29	87.9	325	2 T35271	probable transcrip
31	29	87.9	359	2 T35179	vanillate O-demeth
32	29	87.9	371	1 HUBP1A	hyaluronoglucosam
33	29	87.9	371	2 B39625	T-cell receptor al
34	29	87.9	397	2 A39665	lymphoid enhancer
35	29	87.9	399	2 A39625	T-cell receptor al
36	29	87.9	412	1 A39625	13-methyl-2-oxobut
37	29	87.9	412	2 C72548	probable dihydrotri
38	29	87.9	460	2 A72009	fumurate hydratase
39	29	87.9	460	2 B86617	fumurate hydratase
40	29	87.9	461	2 E71672	fumurate hydratase
41	29	87.9	463	2 B81725	fumurate hydratase
42	29	87.9	463	2 D97826	fumurate hydratase
43	29	87.9	463	2 H71462	probable fumurate
44	29	87.9	463	2 D87510	fumurate hydratase
45	29	87.9	464	2 H83538	fumurate hydratase

## ALIGNMENTS

RESULT 1  
AF2809  
hypothetical protein Atut1896 [imported] - Agrobacterium tumefaciens (strain C58, Dupon  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #ext\_change 09-Jul-2004  
C/Accession: AF2809  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
erage, G.; Gillet, W.; Grant, C.; Guentherer, D.; Kutyavyn, T.; Levy, R.; Li, M.; McCle  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:1608550; PMID:11743193  
A/Accession: AF2809  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-68 <KUR>  
A/Cross-references: UNIPROT:Q8UE65; GB:AE008688; PIDN:AAL42892.1; PID:G17740345; GSPDB  
A/Experimental source: strain C58 (Dupont)  
A/Genetics:  
A/Gene: Atut1896  
A/Map position: circular chromosome

Query Match  
Best local Similarity 100.0%; Score 33; DB 2; Length 68;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 LGTGR 6  
DB 36 LGTGR 41

RESULT 2  
T49710  
related to glucan 1,4-alpha-glucosidase [imported] - Neurospora crassa  
N/Alternate names: protein B23121.230  
C/Species: Neurospora crassa  
C/Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #ext\_change 09-Jul-2004  
C/Accession: T49710  
R/Schulze, U.; Aign, V.; Hehnel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura  
submitted to the Protein Sequence Database, May 2000  
A/Reference number: Z25022  
A/Accession: T49710  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-931 <SCH>  
A/Cross-references: UNIPROT:Q9P5K6; EMBL:AL356172; GSPDB:GN00116; NCSP:B23121.230  
A/Experimental source: BAC clone B23121; strain OR74A  
A/Genetics:  
A/Gene: NCSP:B23121.230  
A/Map position: 6

A:introns: 503/2

Query Match 100.0%; Score 33; DB 2; Length 931;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 797 LGTGR 802

RESULT 3

probable peptidoglycan bound protein (LPXTG motif) lmo0435 [imported] - *Listeria monocytogenes*  
C/Species: *Listeria monocytogenes*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004

C/Accession: A01129  
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entlian, K.D.; Fahl, H.; D.; Jones, L.M.; Karet, U.  
Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kuster, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schuener, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A./Title: Comparative genomics of *Listeria* species.  
A/Reference number: A01077; MUID:21537279; PMID:11679669  
A/Accession: A01129  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-2013 <GUA>  
A/Cross-references: UNIPROT:Q8Y9T8; GB:NC\_003210; PIDN:CA038514.1; PID:G1409812; GSPDB:  
A/Experimental source: strain EGD-e  
C/Genetics:

A:Gene: lmo0435

Query Match 100.0%; Score 33; DB 2; Length 2013;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 892 LGTGR 897

RESULT 4

probable peptidoglycan bound protein (LPXTG motif) lmo0457 [imported] - *Listeria innocua*  
C/Species: *Listeria innocua*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004

C/Accession: A01489  
R/Glaeser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entlian, K.D.; Fahl, H.; D.; Jones, L.M.; Karet, U.  
Science 294, 849-852, 2001  
A:Authors: Kretz, J.; Kuhn, M.; Kuster, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schuener, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A./Title: Comparative genomics of *Listeria* species.  
A/Reference number: A01077; MUID:21537279; PMID:11679669  
A/Accession: A01489  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-2013 <GUA>  
A/Cross-references: UNIPROT:Q92EK2; GB:AL592022; PIDN:CA036589.1; PID:G16412898; GSPDB:G  
A/Experimental source: strain C13D11262  
C/Genetics:

A:Gene: lmo0457

Query Match 100.0%; Score 33; DB 2; Length 2013;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 892 LGTGR 897

RESULT 5

conserved hypothetical protein CC0742 [imported] - *Caulobacter crescentus*  
C/Species: *Caulobacter crescentus*  
C/Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004

C/Accession: C87341  
R/Nierman, W.C.; Feldblym, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A/Title: Complete genome sequence of *Caulobacter crescentus*.  
A/Reference number: A87249; MUID:21173698; PMID:11259647  
A/Accession: C87341  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-188 <STG>  
A/Cross-references: UNIPROT:Q9AA64; GB:AE005673; MID:G13421975; PIDN:AAK22727.1; GSPDB:G  
C/Genetics:

A:Gene: CC0742

Query Match 93.9%; Score 31; DB 2; Length 188;  
Best Local Similarity 83.3%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 130 MGTGR 135

RESULT 6

hypothetical protein GGI-83 [imported] - human  
C/Species: *Homo sapiens* (man)  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004

C/Accession: T44603  
R/Lin, W.C.  
Submitted to the EMBL Data Library, May 1999  
A/Description: Comparative gene cloning: Identification of novel human genes with Caeno  
A/Reference number: 222808  
A/Accession: T44603  
A/Status: preliminary; translated from GB/EMBL/DBD  
A/Molecule type: mRNA  
A/Residues: 1-288 <LIN>  
A/Cross-references: UNIPROT:Q9Y392; EMBL:AF151841; PIDN:AA034078.1  
C/Genetics:

A/Map position: 8

Query Match 90.9%; Score 30; DB 2; Length 288;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
DB 36 VGTGR 41

RESULT 7

probable guanylate kinase [imported] - *Arabidopsis thaliana*  
C/Species: *Arabidopsis thaliana* (mouse-ear cress)  
C/Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004

C/Accession: C84847  
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentol, M.L.; Town, C.D.; Fujii, C.Y.  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umeyama, L.; Tallon,  
ues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: C84847  
A/Status: preliminary  
A/Molecule type: DNA



A;Residues: 1-329 <STO>  
 A;Cross-references: UNIPROT:P93757; GB:AE002093; NID:g6598818; PIDN:AAF18683.1; GSPDB:GN  
 C;Genetics:  
 A;Gene: At2G41880  
 A;Map position: 2

Query Match 90.9%; Score 30; DB 2; Length 329;  
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
 DB 71 LGTGR 76

# RESULT 8

T50675  
 guanylate kinase (EC 2.7.4.8) [imported] - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 09-Jul-2004  
 C;Accession: T50675  
 R;Kumar, V.; Spangenberg, O.; Konrad, M.  
 Eur. J. Biochem. 267, 606-615, 2000  
 A;Title: Cloning of the guanylate kinase homologues AGK-1 and AGK-2 from Arabidopsis th  
 A;Reference number: Z25173; MUID:20098538; PMID:10632732  
 A;Accession: T50675  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-387 <KUM>  
 A;Cross-references: UNIPROT:Q9M681; EMBL:AF204677; PIDN:AAF60252.1  
 A;Experimental source: cultivar Columbia  
 C;Genetics:  
 A;Gene: AGK-1  
 A;Introns: 1/3; 39/3; 65/2; 108/3; 229/3; 315/3; 331/1; 361/2  
 C;Keywords: phosphotransferase

Query Match 90.9%; Score 30; DB 2; Length 387;  
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
 DB 71 LGTGR 76

# RESULT 9

T46191  
 guanylate kinase-like protein - Arabidopsis thaliana  
 N;Alternate names: protein T8H10.150  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
 C;Accession: T46191  
 R;Benes, V.; Rechmann, S.; Bokkova, D.; Ansojge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.  
 submitted to the Protein Sequence Database, January 2000  
 A;Reference number: Z23014  
 A;Accession: T46191  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-391 <BEN>  
 A;Cross-references: UNIPROT:Q9SCB8; EMBL:AL133248  
 A;Experimental source: cultivar Columbia; BAC clone T8H10  
 C;Genetics:  
 A;Map position: 3  
 A;Introns: 4/1; 40/3; 66/2; 109/3; 230/3; 271/2; 316/3; 332/1; 364/2  
 A;Note: T8H10.150

Query Match 90.9%; Score 30; DB 2; Length 391;  
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
 DB 72 LGTGR 77

# RESULT 10

S49980  
 glutamate-5-semialdehyde dehydrogenase (EC 1.2.1.41) - Corynebacterium glutamicum (fr.  
 N;Alternate names: gamma-glutamyl phosphate reductase  
 C;Species: Corynebacterium glutamicum  
 C;Date: 13-Jan-1995 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
 C;Accession: S49980  
 R;Seberjnski, I.; Wojcik, F.; Reyes, O.; Leblon, G.  
 submitted to the EMBL Data Library, November 1994  
 A;Description: Two loci of Corynebacterium glutamicum ATCC17965 that complement Esche  
 A;Reference number: S49977  
 A;Accession: S49980  
 A;Molecule type: DNA  
 A;Residues: 1-432 <SER>  
 A;Cross-references: UNIPROT:P45638; EMBL:X82929; NID:g599719; PIDN:CAA58103.1; PID:g5

Query Match 90.9%; Score 30; DB 2; Length 432;  
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
 DB 36 LGTGR 41

# RESULT 11

A41054  
 fasciclin II, transmembrane splice form precursor - fruit fly (Drosophila melanogaster)  
 C;Species: Drosophila melanogaster  
 C;Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 09-Jul-2004  
 C;Accession: A41054  
 R;Greeningloh, G.; Rehm, E.J.; Goodman, C.S.  
 Cell 67, 45-57, 1991  
 A;Title: Genetic analysis of growth cone guidance in Drosophila: fasciclin II function  
 A;Reference number: A41054; MUID:92005695; PMID:1913818  
 A;Accession: A41054  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-811 <GR2>  
 A;Cross-references: UNIPROT:P34082; GB:M77165; NID:g157402; PID:g157403  
 C;Genetics:  
 A;Gene: FlyBase:FBsg0000635  
 A;Cross-references: FlyBase:FBgn0000635  
 C;Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; int  
 C;Keywords: membrane protein

Query Match 90.9%; Score 30; DB 2; Length 811;  
 Best Local Similarity 83.3%; Pred. No. 3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
 DB 481 VGTGR 486

# RESULT 12

B41054  
 fasciclin II PI-linked splice form precursor - fruit fly (Drosophila melanogaster)  
 C;Species: Drosophila melanogaster  
 C;Date: 21-Apr-1992 #sequence\_revision 21-Apr-1992 #text\_change 17-Mar-2000  
 C;Accession: B41054  
 R;Greeningloh, G.; Rehm, E.J.; Goodman, C.S.  
 Cell 67, 45-57, 1991  
 A;Title: Genetic analysis of growth cone guidance in Drosophila: fasciclin II functions  
 A;Reference number: A41054; MUID:92005695; PMID:1913818  
 A;Accession: B41054  
 A;Status: preliminary

A:Molecule type: mRNA  
 A:Residues: 1-873 <GR>  
 A:Cross-references: GB:M77156  
 C:Genetics:  
 A:Gene: FlyBase:FBgn0000635  
 A:Cross-references: FlyBase:FBgn0000635  
 A:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu  
 C:Keywords: transmembrane protein

Query Match 90.9%; Score 30; DB 2; Length 873;  
 Best Local Similarity 83.3%; Pred. No. 3.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LGTSPR 6  
 Db 481 GTGPR 486

## RESULT 13

A36691  
 Ca2+-transporting ATPase (EC 3.6.3.8), sarcoplasmic reticulum - fruit fly (*Drosophila me*  
 C:Species: *Drosophila melanogaster*  
 C>Date: 28-Jun-1991 #sequence\_revision 30-Jan-1993 #text\_change 09-Jul-2004  
 C:Accession: A36691; S07050  
 R:Author: A.; Varadi, A.  
 Biochem. Biophys. Res. Commun. 173, 872-877, 1990  
 A>Title: Molecular cloning and chromosomal localization of a sarco/endoplasmic reticulum  
 A:Reference number: A36691; PMID:91097592; PMID:2148477  
 A:Accession: A36691  
 A:Status: preliminary  
 A:Molecule type: mRNA

A:Residues: 1-1002 <MAG>  
 A:Cross-references: UNIPROT:P22700; GB:M62892; NID:G158415; PIDN:AA00735.1; PID:G158418  
 R:Varadi, A.; Gilmore-Hedder, M.; Benz Jr., E.J.  
 FEBS Lett. 258, 203-207, 1989  
 A>Title: Amplification of the phosphorylation site - ATP-binding site cDNA fragment of  
 A:Reference number: S07049; PMID:9002469; PMID:2557235  
 A:Accession: S07050  
 A:Molecule type: mRNA

A:Residues: 357-513 <VAR>  
 A:Cross-references: EMBL:X17472  
 A>Note: the authors translated the codon CTC for residue 1 as Thr; the sequence shown fo  
 C:Genetics:  
 A:Gene: FlyBase:Ca-P60A  
 A:Cross-references: FlyBase:FBgn0004551  
 C:Function:

A:Description: catalyzes active transport of Ca2+ ions across cellular membranes; Ca2+  
 C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain  
 C:Keywords: ATP; calcium binding; hydrolase; phosphoprotein; transmembrane protein  
 F:40-57/Domain: calcium binding #status predicted <CA1>  
 F:60-78/Domain: transmembrane #status predicted <TM01>  
 F:87-107/Domain: transmembrane #status predicted <TM02>  
 F:108-257/Domain: intracellular #status predicted <INT1>  
 F:111-131/Domain: calcium binding #status predicted <CA2>  
 F:132-238/Domain: transduction #status predicted <TSD>  
 F:258-277/Domain: transmembrane #status predicted <TM03>  
 F:288-307/Domain: transmembrane #status predicted <TM04>  
 F:308-760/Domain: intracellular #status predicted <INT2>  
 F:310-329/Domain: calcium binding #status predicted <CA3>  
 F:330-505/Domain: catalytic #status predicted <PHY>  
 F:506-680/Domain: ATP binding #status predicted <ATP>  
 F:595-768/Domain: ATPase nucleotide-binding domain homology <ATN>  
 F:763-784/Domain: transmembrane #status predicted <TM05>  
 F:789-809/Domain: transmembrane #status predicted <TM06>  
 F:837-857/Domain: transmembrane #status predicted <TM07>  
 F:894-913/Domain: transmembrane #status predicted <TM08>  
 F:931-950/Domain: transmembrane #status predicted <TM09>  
 F:959-980/Domain: transmembrane #status predicted <TM10>  
 F:951/Active site: Asp (Aspartylphosphate intermediate) #status predicted  
 F:515/Binding site: ATP (Lys) #status predicted

Query Match 90.9%; Score 30; DB 2; Length 1002;  
 Best Local Similarity 83.3%; Pred. No. 3.8e+02;

Qy 2 GTGPR 6  
 Db 70 GTGPR 74

Search completed: November 2, 2004, 20:22:15  
 Job time: 3.10701 secs

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 LGTSPR 6  
 Db 506 LGTSPR 511

## RESULT 14

G70824  
 hypothetical protein RV0748 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: *Mycobacterium tuberculosis*  
 C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
 C:Accession: G70824  
 R:Author: S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 Rajandream, N.A.; Rogers, J.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
 A:Reference number: A70500; PMID:98295987; PMID:9634230  
 A:Accession: G70824  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA

A:Residues: 185 <COL>  
 A:Cross-references: UNIPROT:Q53811; GB:AL021958; GB:AL123456; NID:G3261536; PIDN:CAA175  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0748

Query Match 87.9%; Score 29; DB 2; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GTGPR 6  
 Db 56 GTGPR 60

RESULT 15  
 B95105  
 conserved hypothetical protein SP0910 [imported] - *Streptococcus pneumoniae* (strain TIC  
 C:Species: *Streptococcus pneumoniae*  
 C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
 C:Accession: B95105  
 R:Teitelin, R.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Petersen, S.; Hen  
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple  
 nson, T.; Hickey, E.X.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison  
 A>Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
 A:Reference number: A95000; PMID:21357209; PMID:11463916  
 A:Accession: B95105  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-90 <KUR>  
 A:Cross-references: UNIPROT:Q978B4; GB:AE005672; PIDN:AAK75035.1; PID:G14972384; GSPDB  
 A:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SP0910

Query Match 87.9%; Score 29; DB 2; Length 90;  
 Best Local Similarity 100.0%; Pred. No. 52;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GTGPR 6  
 Db 70 GTGPR 74

Search completed: November 2, 2004, 20:22:15  
 Job time: 3.10701 secs

Sat Nov 6 18:59:22 2004

us-10-054-873-3.rapb

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 2, 2004, 20:20:47; Search time 4.4946 Seconds  
(without alignments)  
432.820 Million cell updates/sec

Title: US-10-054-873-3

Perfect score: 33

Sequence: 1 LGTPR 6

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1370721 seqs, 324215800 residues

Total number of hits satisfying chosen parameters: 1370721

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications\_AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	6	US-10-054-873-3	Sequence 290, App
2	33	100.0	15	US-10-337-598-290	Sequence 3, Appl
3	33	100.0	82	US-10-437-963-168327	Sequence 168327,
4	33	100.0	107	US-10-054-873-6	Sequence 6, Appl
5	33	100.0	135	US-10-408-765A-1525	Sequence 1525, Ap
6	33	100.0	150	US-10-054-873-7	Sequence 7, Appl
7	33	100.0	275	US-10-369-493-12819	Sequence 12819, A
8	33	100.0	802	US-10-437-963-178776	Sequence 178776,
9	33	100.0	952	US-10-108-260A-3284	Sequence 3284, Ap
10	33	100.0	2013	US-10-282-122A-60608	Sequence 60608, A
11	33	100.0	7257	US-10-014-717-5	Sequence 5, Appl
12	31	93.9	455	US-10-437-963-154712	Sequence 154712,
13	31	93.9	1963	US-10-437-963-155445	Sequence 155445,

14	30	90.9	24	US-09-864-761-39589	Sequence 39589, A
15	30	90.9	78	US-10-437-963-188463	Sequence 188463,
16	30	90.9	78	US-10-767-701-56946	Sequence 56946, A
17	30	90.9	88	US-10-424-599-183036	Sequence 183036,
18	30	90.9	114	US-09-728-626-3668	Sequence 3668, Ap
19	30	90.9	186	US-10-767-701-36431	Sequence 36431, A
20	30	90.9	208	US-10-233-553-30	Sequence 30, Appl
21	30	90.9	222	US-10-240-145-161	Sequence 161, App
22	30	90.9	231	US-10-243-552-564	Sequence 564, App
23	30	90.9	268	US-10-424-599-14755	Sequence 14755,
24	30	90.9	274	US-10-424-599-14755	Sequence 14755,
25	30	90.9	275	US-10-425-114-48504	Sequence 48504, A
26	30	90.9	298	US-09-925-301-1224	Sequence 1224, Ap
27	30	90.9	298	US-10-296-115-1307	Sequence 1307, Ap
28	30	90.9	309	US-10-425-114-47615	Sequence 47615, A
29	30	90.9	309	US-10-425-114-48066	Sequence 48066, A
30	30	90.9	377	US-10-437-963-173919	Sequence 173919,
31	30	90.9	432	US-09-728-626-6081	Sequence 6081, Ap
32	30	90.9	471	US-10-029-386-31556	Sequence 31556, A
33	30	90.9	575	US-10-425-114-64569	Sequence 64569, A
34	30	90.9	575	US-10-425-114-64569	Sequence 64569, A
35	30	90.9	676	US-10-374-979-101	Sequence 101, App
36	30	90.9	676	US-10-182-936A-101	Sequence 101, App
37	30	90.9	804	US-10-128-714-3088	Sequence 3088, Ap
38	30	90.9	1028	US-10-128-714-8098	Sequence 8098, Ap
39	30	90.9	1148	US-10-437-963-188462	Sequence 188462,
40	30	90.9	1204	US-10-437-963-188471	Sequence 188471,
41	30	90.9	1233	US-10-437-963-188471	Sequence 188471,
42	30	90.9	1467	US-10-437-963-186949	Sequence 186949,
43	30	90.9	1540	US-10-437-963-186461	Sequence 186461,
44	30	90.9	2071	US-10-437-963-103806	Sequence 103806,
45	30	90.9	2978	US-10-437-963-121042	Sequence 121042,

#### ALIGNMENTS

RESULT 1  
US-10-054-873-3  
; Sequence 3, Application US/10054873  
; Publication No. US20020164712A1  
GENERAL INFORMATION:  
APPLICANT: Gan, Zhong Ru  
TITLE OF INVENTION: Chimeric Protein Containing an  
Intermolecular Chapterone-Like Sequence  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESS: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor,  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/054,873  
FILING DATE: 22-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/CN98/00052  
FILING DATE: 31-MAR-1998  
APPLICATION NUMBER: US 09/423,100  
FILING DATE: 11-DEC-2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Myciotek, Frank J  
REGISTRATION NUMBER: 46,946  
REFERENCE/DOCKET NUMBER: 020167-000330US  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

LENGTH: 6 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 3  
US-10-054-873-3

Query Match 100.0%; Score 33; DB 13; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.2e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 1 LGTGR 6

## RESULT 2

US-10-327-598-290  
Sequence 280 Application US/10327598  
Publication No. US20040181039A1

GENERAL INFORMATION:  
APPLICANT: Krah, Eugene  
APPLICANT: Guo, Hongliang  
APPLICANT: Aiyappa, Ashok  
APPLICANT: Lawton, Robert  
TITLE OF INVENTION: Canine Immunoglobulin Variable Domains, Caninized Antibodies, and  
FILE REFERENCE: 01-799-A  
CURRENT APPLICATION NUMBER: US/10/327,598  
CURRENT FILING DATE: 2002-12-20  
PRIOR APPLICATION NUMBER: US 60/344,874  
PRIOR FILING DATE: 2001-12-21  
NUMBER OF SEQ ID NOS: 1139  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 290  
LENGTH: 15  
TYPE: prt  
ORGANISM: canis familiaris;

Query Match 100.0%; Score 33; DB 16; Length 15;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 6 LGTGR 11

## RESULT 3

US-10-437-963-168327  
Sequence 168327 Application US/10437963  
Publication No. US20040123343A1

GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barzduk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 168327  
LENGTH: 82  
TYPE: prt  
ORGANISM: Oryza sativa

FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(82)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_66852C.1.pcp  
US-10-437-963-168327

Query Match 100.0%; Score 33; DB 16; Length 82;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 13 LGTGR 18

## RESULT 4

US-10-054-873-6  
Sequence 6 Application US/10054873  
Publication No. US20020164712A1

GENERAL INFORMATION:  
APPLICANT: Gan, Zhong Ru  
TITLE OF INVENTION: Chimeric Protein Containing an  
Intramolecular Chaperone-Like Sequence  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/054,873  
FILING DATE: 22-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/CN98/00052  
FILING DATE: 31-MAR-1998  
APPLICATION NUMBER: US 09/423,100  
FILING DATE: 11-DEC-2000

ATTORNEY/AGENT INFORMATION:  
NAME: Mycroft, Frank J  
REGISTRATION NUMBER: 46,946  
REFERENCE/DOCKET NUMBER: 020167-000130US  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 107 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6  
US-10-054-873-6

Query Match 100.0%; Score 33; DB 13; Length 107;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 50 LGTGR 55

## RESULT 5

US-10-408-765A-1525  
Sequence 1525 Application US/10408765A

Publication No. US20040101874A1  
GENERAL INFORMATION:  
APPLICANT: Ghosh, Soumitra S.  
APPLICANT: Fahy, Boi D.  
APPLICANT: Zhang, Bing  
APPLICANT: Gibson, Bradford W.  
APPLICANT: Taylor, Steven W.  
APPLICANT: Wernick, Gary M.  
APPLICANT: Wernick, Dale E.  
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
FILE REFERENCE: 660088.465  
CURRENT APPLICATION NUMBER: US/10/408,765A  
CURRENT FILING DATE: 2003-04-04  
NUMBER OF SEQ ID NOS: 3077  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1525  
LENGTH: 135  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-408-765A-1525

Query Match 100.0%; Score 33; DB 16; Length 135;  
Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 5 LGTGR 10

RESULT 6  
US-10-054-873-7  
Sequence 7, Application US/10054873  
Publication No. US20020164712A1  
GENERAL INFORMATION:  
APPLICANT: Gan, Zhong Ru  
TITLE OF INVENTION: Chimeric Protein Containing an  
Intramolecular Chaperone-Like Sequence  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/054,873  
FILING DATE: 22-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/CN98/00052  
FILING DATE: 31-MAR-1998  
APPLICATION NUMBER: US 09/423,100  
FILING DATE: 11-DEC-2000  
ATTORNEY/AGENT INFORMATION:  
NAME: Mycroft, Frank J  
REGISTRATION NUMBER: 46,946  
REFERENCE/DOCKET NUMBER: 020167-000130US  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 150 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-10-054-873-7  
Query Match 100.0%; Score 33; DB 13; Length 150;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 93 LGTGR 98

RESULT 7  
US-10-369-493-12819  
Sequence 12819, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 12819  
LENGTH: 275  
TYPE: PRT  
ORGANISM: Aspergillus nidulans  
US-10-369-493-12819

Query Match 100.0%; Score 33; DB 14; Length 275;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
Db 15 LGTGR 20

RESULT 8  
US-10-437-963-178776  
Sequence 178776, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovacic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukhartov, Andrey A.  
APPLICANT: Barbaruk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 178776  
LENGTH: 802  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_7629C.1.pep  
US-10-437-963-178776

Query Match 100.0%; Score 33; DB 16; Length 802;  
Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 LGTGR 6  
|||||  
DB 279 LGTGR 284

RESULT 9  
US-10-108-260A-3284  
Sequence 3284, Application US/10108260A  
Publication No. US20040005560A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20040005560A1 full length cDNA  
FILE REFERENCE: H1-A0106  
CURRENT APPLICATION NUMBER: US/10/108,260A  
CURRENT FILING DATE: 2002-03-27  
NUMBER OF SEQ ID NOS: 5458  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3284  
LENGTH: 952  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-108-260A-3284

Query Match 100.0%; Score 33; DB 15; Length 952;  
Best Local Similarity 100.0%; Pred. No. 9,8e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
DB 634 LGTGR 639

RESULT 10  
US-10-282-122A-60608  
Sequence 60608, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PAM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 60608  
LENGTH: 2013  
TYPE: PRT  
ORGANISM: Listeria monocytogenes  
US-10-282-122A-60608

Query Match 100.0%; Score 33; DB 15; Length 2013;  
Best Local Similarity 100.0%; Pred. No. 2e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
DB 892 LGTGR 897

RESULT 11  
US-10-014-717-5  
Sequence 5, Application US/10014717  
Publication No. US20020192778A1  
GENERAL INFORMATION:  
APPLICANT: Schupp, Thomas  
APPLICANT: Ligon, James  
APPLICANT: Molnar, Istvan  
APPLICANT: Zirkle, Ross  
APPLICANT: Cyr, Devon  
APPLICANT: Goerlach, Joern  
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES  
FILE REFERENCE: 4-10582A  
CURRENT APPLICATION NUMBER: US/10/014,717  
CURRENT FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US/09/335,409  
PRIOR FILING DATE: 1999-06-17  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 7257  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-10-014-717-5

Query Match 100.0%; Score 33; DB 13; Length 7257;  
Best Local Similarity 100.0%; Pred. No. 6,5e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6  
|||||  
DB 1041 LGTGR 1046

RESULT 12  
US-10-437-963-154712  
Sequence 154712, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovacic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Bouharov, Audrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966

```

; SEQ ID NO 154712
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(455)
; OTHER INFORMATION: unsure at all Xaa locations
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54545C.1.pcp
US-10-437-963-154712

Query Match
Best Local Similarity 93.3%; Score 31; DB 16; Length 455;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6
Db 336 MGTGR 343

RESULT 13
US-10-437-963-155445
; Sequence 155445, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 155445
; LENGTH: 1963
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55207C.1.pcp
US-10-437-963-155445

Query Match
Best Local Similarity 93.3%; Score 31; DB 16; Length 1963;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6
Db 336 MGTGR 341

RESULT 14
US-09-864-761-39589
; Sequence 39589, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
```

```

; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anncmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39589
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004061.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
US-09-864-761-39589

Query Match
Best Local Similarity 90.3%; Score 30; DB 9; Length 24;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6
Db 13 LGTGR 18

RESULT 15
US-10-437-963-188463
; Sequence 188463, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
```

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21 (53221) B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ. ID NOS: 204966  
SEQ. ID NO 188463  
LENGTH: 78  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_85064C.1.pap  
US-10-437-963-188463

Query Match 90.9%; Score 30; DB 16; Length 78;  
Best Local Similarity 83.3%; Pred. No. 3.4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CY 1 LGTGPR 6  
:|||||  
Db 30 VGTGPR 35

Search completed: November 2, 2004, 20:59:21  
Job time : 6.49446 secs



Sat Nov 6 18:59:23 2004

us-10-054-873-3.rup

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 2, 2004, 19:48:36 ; Search time 5.6181 Seconds  
(without alignments)  
594.006 Million cell updates/sec

Title: US-10-054-873-3  
Perfect score: 33  
Sequence: 1 LGTGR 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 02.\*  
1: uniprot\_sprot.\*  
2: uniprot\_tramb1.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	68	Q8UE65	Q8UE65 agrobacteri
2	33	100.0	135	Q8HA29	Q8HA29 homo sapien
3	33	100.0	277	Q8L1U5	Q8L1U5 bordetella
4	33	100.0	287	Q7W361	Q7W361 bordetella
5	33	100.0	287	Q7W361	Q7W361 bordetella
6	33	100.0	289	Q7W361	Q7W361 bordetella
7	33	100.0	491	Q8P623	Q8P623 xanthomonas
8	33	100.0	563	Q810F1	Q810F1 drosophila
9	33	100.0	585	Q8HEG7	Q8HEG7 homo sapien
10	33	100.0	736	Q6YBM4	Q6YBM4 cryptococcus
11	33	100.0	736	AA025635	AA025635 cryptococcus
12	33	100.0	896	Q6R8A4	Q6R8A4 homo sapien
13	33	100.0	896	Q6R8A4	Q6R8A4 homo sapien
14	33	100.0	946	Q8P5K6	Q8P5K6 neurospora
15	33	100.0	1095	Q8KLS0	Q8KLS0 rhodobacter
16	33	100.0	1522	Q15069	Q15069 homo sapien
17	33	100.0	2013	Q92EK2	Q92EK2 listeria in
18	33	100.0	2013	Q8Y9T8	Q8Y9T8 listeria in
19	33	100.0	7257	Q91AC7	Q91AC7 polyanthum
20	31	93.9	188	Q8R3A4	Q8R3A4 caulobacter
21	31	93.9	297	Q8K0S5	Q8K0S5 methylodact
22	31	93.9	376	Q7W0F9	Q7W0F9 rhodospirillum
23	31	93.9	448	Q8G430	Q8G430 bifidobacter
24	31	93.9	485	Q6YX20	Q6YX20 oryza sativ
25	31	93.9	485	BAD10610	BAD10610 oryza sat
26	31	93.9	541	Q6ZFA9	Q6ZFA9 oryza sativ
27	31	93.9	541	BAD17010	BAD17010 oryza sat
28	31	93.9	543	Q6ZFA9	Q6ZFA9 oryza sativ
29	31	93.9	543	BAD09254	BAD09254 oryza sat
30	31	93.9	561	Q9DK04	Q9DK04 alipahuayo
31	31	93.9	579	Q9LDD30	Q9LDD30 cryptocodi

32	31	93.9	656	Q96529	Q96529 meloidogyne
33	31	93.9	656	Q9XVA9	Q9XVA9 meloidogyne
34	31	93.9	676	Q6ZFB7	Q6ZFB7 oryza sativ
35	31	93.9	676	BAD17002	BAD17002 oryza sat
36	31	93.9	688	Q6ZFT2	Q6ZFT2 oryza sativ
37	31	93.9	688	BAD09243	BAD09243 oryza sat
38	30	90.9	114	Q8NU02	Q8NU02 corynebacte
39	30	90.9	114	CAF18716	CAF18716 corynebacte
40	30	90.9	128	Q8FU73	Q8FU73 corynebacte
41	30	90.9	151	Q93WV9	Q93WV9 musa acumin
42	30	90.9	196	Q6MPF9	Q6MPF9 bdellovibrio
43	30	90.9	196	CAE78708	CAE78708 bdellovib
44	30	90.9	200	Q8KXW0	Q8KXW0 rubegeia sp
45	30	90.9	209	Q9H9X7	Q9H9X7 homo sapien

#### ALIGNMENTS

RESULT 1	Q8UE65	PRELIMINARY:	PRT:	68 AA.
AC	Q8UE65;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DE	Hypothetical protein Atcu1896.			
GN	OrderedLocustNames=Atcu1896;			
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.			
OX	NCBI_TaxID=176299;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Dupont;			
RX	MEDLINE=21608550; PubMed=11741193;			
RA	Wood D.W., Sebubal J.C., Kaul R., Monks D.B., Kiteajima J.P.,			
RA	Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,			
RA	Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boye D., Sr.,			
RA	Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,			
RA	Kutyavina T., Levy R., Li M.-J., McClelland E., Palmeri A.,			
RA	Raymond C., Rouse G., Saephammachak C., Wu Z., Romero P., Gordon D.,			
RA	Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,			
RA	Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,			
RA	Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,			
RT	Nester E.W.,			
RT	"The genome of the natural genetic engineer Agrobacterium tumefaciens			
RL	C58."			
RL	Science 294:2317-2323(2001).			
DR	EMBL; AB009143; AAL42892.1; -.			
DR	PIR; AF2809; AF2809.			
KW	Complete proteome; Hypothetical protein.			
SC	SEQUENCE 68 AA; 7596 MW; AEGCBDB94613949F CRC64;			
QY	Query Match	100.0%;	Score 33;	DB 2; Length 68;
DB	Best Local Similarity	100.0%;	Pred. No. 47;	
	Matches 6; Conservative	0;	Mismatches	0; Gaps 0;
			Indels	0;
RESULT 2	Q9HA29	PRELIMINARY:	PRT;	135 AA.
AC	Q9HA29;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Hypothetical protein FLJ12345.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			

```

RX MEDLINE=22215669; PubMed=12228265;
RA Murphy E.R., Sacco R.E., Dickenson A., Metzger D.J., Hu Y.,
DR GO: GO:0005381; P:iron ion transporter activity; IEA.
RT "Onoroff P.E., Connell T.D.;
RT "Mun, a virulence-associated outer membrane protein of Bordetella
RT avium, is required for the acquisition of iron from heme and
RL hemoproteins." 70:5390-5403 (2002).
RL Infect. Immun. 70:5390-5403 (2002).
DR EMBL: AY095952; AAM28270.1; -.
DR GO: GO:0005381; P:iron ion transporter activity; IEA.
DR GO: GO:0006827; P:high affinity iron ion transport; IEA.
DR InterPro: IPR002491; Peripla_BP.
DR Pfam: PF01497; Peripla_BP_2; 1.
SQ SEQUENCE 277 AA; 26898 MW; 7FCDCCDCA37B4D CRC64;
SQ

Oy 1 LGTGP 6
Db 252 LGTGP 257

Query Match 100.0%; Score 33; DB 2; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LGTGP 6
Db 252 LGTGP 257

RESULT 4
O7W361 PRELIMINARY; PRT; 287 AA.
ID O7W361;
AC O7W361;
DT 01-OCT-2003 (TREMBLrel. 25; Created)
DT 01-OCT-2003 (TREMBLrel. 25; Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26; Last annotation update)
DE Putative hemin binding protein.
GN Name=Ohr1; OrderedLocustNames=BP187;
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
CX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=12822 / ATCC BAA-567;
RA MEDLINE=22627954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parikhil J., Sebathia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerceno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achtman M., Actin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Collingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtwood S., Jagals K.,
RA Leather S., Mole S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds N., Skelton J., Squares R., Squares S., Stevens K.,
RA Urwin L., Whitehead S., Barrett B.G., Maxwell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40 (2003).
DR EMBL: BG640436; CAB39466.1; -.
DR GO: GO:0005381; P:iron ion transporter activity; IEA.
DR GO: GO:0006827; P:high affinity iron ion transport; IEA.
DR InterPro: IPR002491; Peripla_BP.
DR Pfam: PF01497; Peripla_BP_2; 1.
KM Complete proteome.
SQ SEQUENCE 287 AA; 29393 MW; 86F8317AD5241C40 CRC64;
SQ

Query Match 100.0%; Score 33; DB 2; Length 287;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 LGTGP 6
Db 259 LGTGP 264

RESULT 5
O7W368 PRELIMINARY; PRT; 287 AA.
ID O7W368

```

AC Q7MEH8; 07MEH8; 25. Created)

DT 01-OCT-2003 (TREMBlrel. 25, last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, last annotation update)

DE Putative hemin binding protein.

GN Name=hnt; OrderedLocNames=BB4657;

OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

CC Alcaligenaceae; Bordetella.

OK NCBI\_TaxID=518;

RN 111

RP SEQUENCE FROM N.A. ATCC BAA-588;

RC STRAIN=RB50 / ATCC BAA-588;

RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;

RA Parthill J., Sebailia M., Preston A., Murphy L.D., Thomson N.R.,

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,

RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,

RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,

RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,

RA Leather S., Moulé S., Norberczak H., O'Neill S., Ormond D., Price C.,

RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,

RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,

RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.,

RT "Comparative analysis of the genome sequences of Bordetella pertussis,

RT Bordetella parapertussis and Bordetella bronchiseptica."

RL Nat. Genet. 35:32-40(2003).

DR GO; GO:0005381; F:iron ion transporter activity; IEA.

DR GO; GO:0006827; P:high affinity iron ion transport; IEA.

DR InterPro; IPR002491; Peripla\_BP.

DR Pfam; PF01497; Peripla\_BP\_2; 1.

KW Complete proteome.

SQ SEQUENCE 287 AA; 29363 MW; 96F9317AC5251031 CRC64;

QY Query Match 100.0%; Score 33; DB 2; Length 287;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 LGTGR 6

259 LGTGR 264

RESULT 6

ID Q7VSO6; PRELIMINARY; PRT; 269 AA.

AC Q7VSO6; PRELIMINARY; PRT; 269 AA.

DT 01-OCT-2003 (TREMBlrel. 25, last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, last annotation update)

DE Putative hemin binding protein.

GN Name=hnt; OrderedLocNames=BB4657;

OS Bordetella pertussis.

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

CC Alcaligenaceae; Bordetella.

OK NCBI\_TaxID=520;

RN 111

RP SEQUENCE FROM N.A. ATCC BAA-589 / NCTC 13251;

RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;

RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;

RA Parthill J., Sebailia M., Preston A., Murphy L.D., Thomson N.R.,

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,

RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,

RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,

RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,

RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,

RA Leather S., Moulé S., Norberczak H., O'Neill S., Ormond D., Price C.,

RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,

RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,

RA Unwin L., Whitehead S., Barrett B.G., Maskell D.J.,

RT "Comparative analysis of the genome sequences of Bordetella pertussis,

RT Bordetella parapertussis and Bordetella bronchiseptica."

RL Nat. Genet. 35:32-40(2003).

DR EMBL; BX640412; CA64677.1; -

DR GO; GO:0005381; F:iron ion transporter activity; IEA.

DR GO; GO:0006827; P:high affinity iron ion transport; IEA.

DR InterPro; IPR002491; Peripla\_BP.

DR Pfam; PF01497; Peripla\_BP\_2; 1.

KW Complete proteome.

SQ SEQUENCE 289 AA; 29505 MW; 3B80C28C1D8940AD CRC64;

QY Query Match 100.0%; Score 33; DB 2; Length 289;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 LGTGR 6

261 LGTGR 266

RESULT 7

ID Q8P6Z3; PRELIMINARY; PRT; 491 AA.

AC Q8P6Z3; PRELIMINARY; PRT; 491 AA.

DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, last annotation update)

DE Protein-glutamate methyltransferase.

GN Name=cheB; OrderedLocNames=XCC2822;

OS Xanthomonas campestris (pv. campestris).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

CC Xanthomonadaceae; Xanthomonas.

OK NCBI\_TaxID=3440;

RN 111

RP SEQUENCE FROM N.A. NCPB 528;

RC STRAIN=ATCC 33913 / NCPB 528;

RX MEDLINE=22022145; PubMed=12024217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,

RA Queiroz R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

RA Alves L.M.C., de Amaral A.M., Bertolini M.C., Camargo L.E.A.,

RA Camarotte G., Canavari F., Cardozo J., Chamberg F., Ciapina L.P.,

RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

RA Formigueri E.F., Franco M.C., Greggio C.C., Gruber A., Lemos M.V.F.,

RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Melandis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spicola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,

RA Tindade dos Santos M., Trufi D., Tsai S.M., White F.F.,

RA Secubal J.C., Kitajima J.P.,

RT "Comparison of the genomes of two Xanthomonas pathogens with differing

RT host specificities."

RL Nature 417:459-463(2002).

DR EMBL; AEO12394; AAM42094.1; -

DR HSPD; P04432; ICHD.

DR GO; GO:0004871; F:signal transducer activity; IEA.

DR GO; GO:0006935; P:chemotaxis; IEA.

DR GO; GO:0007165; P:signal transduction; IEA.

DR InterPro; IPR000673; CheB\_methylase.

DR InterPro; IPR011247; Chmx\_methylase.

DR Pfam; PF01339; CheB\_methylase; 1.

DR PIRSF; PIRSF036461; Chmx\_methylase; 1.

DR ProDom; PD003328; CheB\_methylase; 1.

DR PROSITE; PS0123; CheB\_1; 1.

DR PROSITE; PS00290; IG\_MHC; UNKNOWN\_1.

KW Complete proteome.

SQ SEQUENCE 491 AA; 51780 MW; 379E3413A027F619 CRC64;

QY Query Match 100.0%; Score 33; DB 2; Length 491;

Best Local Similarity 100.0%; Pred. No. 3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stachewicz M., Soares M.B., Bonaldo M.F., Casavant J.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Fange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Millaly S.D.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.D., Huiy S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Spevchenko Y., Bouffard G.C.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Small D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP SEQUENCE FROM N.A.  
 RC TISSUE-Testis;  
 RA Straussberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC031069; AA031069.1; -  
 SQ SEQUENCE 585 AA; 66464 MW; 2B2D5F46647D448C CRC64;

Query Match 100.0%; Score 33; DB 2; Length 585;  
 Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTGR 6  
 DB 5 LGTGR 10

RESULT 10  
 ID 06YBM4 PRELIMINARY; PRT; 736 AA.  
 AC 06YBM4;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE TAC3.  
 GN Name=TAC3;  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA PubMed=15207008;  
 RA Still I.H., Veltaiorkumakanauv A.K., DiMatteo A., Liang P.;  
 RT "Structure-function evolution of the transforming acidic coiled coil  
 RT genes revealed by analysis of phylogenetically diverse organisms.";  
 RL BMC Evol. Biol. 4:16-16(2004).  
 DR EMBL; AY161270; AA025635.1; -  
 SQ SEQUENCE 736 AA; 77061 MW; A798FB1C17EF3C8 CRC64;

GN TAC3  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Still I.H., Liang P.;  
 RT "Evolutionary analysis reveals lack of evidence for a fourth TAC3 gene  
 RT member in vertebrates.";  
 RT Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY161270; AA025635.1; -  
 SQ SEQUENCE 736 AA; 77061 MW; A798FB1C17EF3C8 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 736;  
 Best Local Similarity 100.0%; Pred. No. 4.4e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTGR 6  
 DB 207 LGTGR 212

RESULT 12  
 ID 06ZRS4 PRELIMINARY; PRT; 896 AA.  
 AC 06ZRS4;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein FLJ46145.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,  
 RA Matsunoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,  
 RA Matsunura Y., Moriya S., Chiba S., Momiyama H., Onogawa S.,  
 RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Teraoka R.,  
 RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,  
 RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,  
 RA Ishii S., Yamamoto J., Iseno Y., Kawai-Hiro Y., Saito K., Nishikawa T.,  
 RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,  
 RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,  
 RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,  
 RA Sugano S., Nagahara K., Masuko Y., Nagai K., Isegaki T.;  
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK128026; BAC87235.1; -  
 SQ SEQUENCE 896 AA; 98946 MW; DBC9FE06CF7B2C0 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 896;  
 Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 LGTGR 6  
 DB 578 LGTGR 583

RESULT 13  
 ID BAC87235 PRELIMINARY; PRT; 896 AA.  
 AC BAC87235;  
 DT 02-MAR-2004 (TREMBlrel. 27, Created)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
 DE CDNA FLJ46145 f1s, Clone TEST14000370.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohara N.,
RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Matsumura Y., Moriya S., Chida E., Momiyama H., Onogawa S.,
RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Tetsuhara Y.,
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,
RA Iehi S., Yamamoto J., Iseno Y., Kawai-Hito Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Magatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuko Y., Nagai K., Isogai T.,
RT "NEBO human cDNA sequencing project."
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK128026; BAC87235.1; -
SQ SEQUENCE 996 AA; 98946 MW; DBC9BF0B6CF7B2C0 CRC64;

Query Match 100.0%; Score 33; DB 2; Length 996;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6
DB 578 LGTGR 583

RESULT 14
Q9P5K6 PRELIMINARY; PRT; 946 AA.
AC Q9P5K6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Related to glucan 1, 4-alpha-glucosidase.
CN Name=E23L21.230;
OS Neurospora crassa;
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Altm V., Hohnselt J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Neues H.W., Mannhaupt G.,
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Getman Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356172; CAB91691.2; -
DR PIR; T49710; T49710.
SQ SEQUENCE 946 AA; 101461 MW; A8564328338B6E1C CRC64;

Query Match 100.0%; Score 33; DB 2; Length 946;
Best Local Similarity 100.0%; Pred. No. 5.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6
DB 807 LGTGR 812

RESULT 15
Q8KLS0 PRELIMINARY; PRT; 1095 AA.
AC Q8KLS0;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative histidine protein kinase.
GN Name=chea3;

```

```

OS Rhodospirillum rubrum (Rhodospirillum rubrum)
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Rhodospirillum.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS8N;
RA Porter S.L., Warren A.V., Armitage J.P.,
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY488585; CAD32761.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0001603; P:two-component signal transduction system (p. . .); IEA.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR002545; CheW.
DR InterPro; IPR008207; Hct.
DR InterPro; IPR008208; Hct_N.
DR Pfam; PF01584; CheW; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR Pfam; PF01627; Hct; 1.
DR ProDom; PD003142; Hct_N; 1.
DR SMART; SM00260; CheW; 1.
DR SMART; SM00073; HPT; 1.
DR ProSite; PS50894; HPT; 1.
KM kinase
SQ SEQUENCE 1095 AA; 114521 MW; F43CF5A0EBAF3F0E CRC64;

Query Match 100.0%; Score 33; DB 2; Length 1095;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LGTGR 6
DB 1031 LGTGR 1036

```

Search completed: November 2, 2004, 20:20:31  
 Job time: 7.8181 secs